NATIONAL BUREAU OF STANDARDS REPORT

7148

HANDLING OF ADAPTED AND COMPOUND WORDS IN THE NATIONAL BUREAU OF STANDARDS' SCHEME OF MECHANICAL TRANSLATION

by

Owen McArdle

U. S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. Research projects are also performed for other government agencies when the work relates to and supplements the basic program of the Bureau or when the Bureau's unique competence is required. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers. These papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three periodicals available from the Government Printing Office: The Journal of Research, published in four separate sections, presents complete scientific and technical papers; the Technical News Bulletin presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: Monographs, Applied Mathematics Series, Handbooks, Miscellaneous Publications, and Technical Notes.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.25) and its Supplement ($1.50), available from the Superintendent of Documents, Government Printing Office, Washington 25, D.C.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT
1102-40-11513

May 17, 1960

NBS REPORT
7148

HANDLING OF ADAPTED AND COMPOUND WORDS IN THE NATIONAL BUREAU OF STANDARDS' SCHEME OF MECHANICAL TRANSLATION

by

Owen McArdle

Guest Worker
APPLIED MATHEMATICS DIVISION

Technical Report
to
Department of the Army
Office of Ordnance Research and
the Office of the Chief Signal Officer,
Research and Development Division

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

HANDLING OF ADAPTED AND COMPOUND WORDS IN THE NATIONAL BUREAU OF STANDARDS' SCHEME OF MECHANICAL TRANSLATION*

This report is an extension to 1/. It is assumed that the reader is familiar with 1/ and in particular with the terminology used therein.

The linguistic information given herein is based principally on 2/, 3/, and 4/.

---

* This work was sponsored by the Department of the Army, Office of Ordnance Research and the Office of the Chief Signal Officer, Research and Development Division.

PART I. DEFINITIONS

A. Compound Words.

By the term "compound word" we shall denote a word consisting of several Slavic stems, or a hybrid formed by an association of Slavic and non-Slavic elements. The following elements may be found in virtually any combination within a compound word.

a. Slavic root.

b. Slavic prefix.

c. Slavic suffix.

d. Non-Slavic root.

e. Non-Slavic prefix.

f. Non-Slavic suffix.

A compound word combines a base element, usually with an inflectional ending, and one or more attributive elements attached to it in one of the following ways:

a. The original inflectional ending (if any) is replaced by the linking vowel о .

For example:

СВЕТЛЫЙ	+	СЕРЫЙ	becomes	СВЕТЛОСЕРЫЙ
light		gray		light-gray

Some examples of common attributive elements, thus modified are:

БЛАГО	ЗАКОНО	НОВО	РУКО
ВЕРЕТЕНО	ЗВЕЗДО	ОДНО	САМО
ВЗАИМНО	ИГЛО	ПЕРВО	СЕДЛО
ВИДО	КОНЕЧНО	ПОЛНО	СЛОВО
ВЫСОКО	КОСО	ПРОТИВО	СОСРЕДО
ДОЛГО	КРАТКО	ПРЯМО	УМО.
ДРОБНО	КРИВО	РАВНО	ЦЕЛО
ЕСТЕСТВО	МНОГО	РАЗНО	ШАРО

b. The original inflectional ending (if any) of the attributive element is replaced by the linking vowel е .

For example:

ЗЕМЛЯ	+	ТРЯСТИ	becomes	ЗЕМЛЕТРЯСЕНИЕ
earth		to shake		earthquake.

Some examples of attributive elements, formed in this manner are:

ВЫШЕ	ОЧЕ
НИЖЕ	СВЕЖЕ
ОБЩЕ	СВОЕ

ЦЕЛЕ

c. An attributive element is separated from the base element by a hyphen. Such a compound word is often formed by replacing the inflectional ending of the attributive element by a linking vowel followed by a hyphen or dash.

The following combinations may occur:

1) Attributive element Slavic, base element Slavic.

ЮГО - ВОСТОК
ВОГНУТО - ВЫПУКЛЫЙ

2) Attributive element Slavic, base element non-Slavic.

ДРОБНО - ЛИНЕЙНАЯ

3) Attributive element non-Slavic, base element Slavic.

ИНТОНАЦИОННО - СМЫСЛОВОЙ

4) Attributive element non-Slavic, base element non-Slavic.

ФОРМАЛЬНО - ЛОГИЧЕСКИЙ
ЦЕНТРАЛЬНО - СИММЕТРИЧНЫЙ
ИДЕЙНО - ПРИНЦИПИАЛЬНЫЙ

5) Particles, adverbs, and prepositions serving as attributive elements.

КАКИЕ - ЛИБО
КОЕ - КУДА
ИЗ - ЗА

6) At least one element is a proper name.

РИМАН — КОШИ

d. An attributive element is in the genitive case.

For example:

ДВУХ	+	ЛЕТНИЙ	becomes	ДВУХЛЕТНИЙ
two		yearly		two-yearly.

These attributive elements are usually numerals, such as:

ДВУХ ТРЕХ
ПЯТИ ЧЕТЫРЕХ
ШЕСТИ

e. Compound words may consist of complete words without any linking elements:

СВЕРХКРИТИЧЕСКИЙ

МЕЖДУНАРОДНЫЙ

f. Compound words may be formed by ПОЛ and ПОЛУ

1) A hyphen is added --

a) Before a capital letter

ПОЛ - МОСКВЫ

b) Before a vowel

ПОЛ - ОЗЕРА ПОЛ - УЛИЦЫ

c) Before an Л

ПОЛ - ЛИСТА

2) A hyphen is not used in such combinations as:

ПОЛДЕНЬ ПОЛКОМНАТЫ

ПОЛГОДА ПОЛТЕТРАДИ

3) A hyphen is not used in compounds formed with ПОЛУ.

ПОЛУДНЯ ПОЛУКРУГ

ПОЛУСФЕРА

g. Hybrids can occur in the following forms:

1) Non-Slavic element, Slavic root.

АВТОЗАВОД
АГРОПОМОЩЬ
АНТИНАРОДНЫЙ
НИЛЬСТЕПЕННЫЙ
ЦЕНТРОСТРЕМИТЕЛЬНЫЙ

2) Slavic element, non-Slavic root

ВЫСОКОТЕМПЕРАТУРНЫЙ

СРЕДНЕКВАДРАТИЧНЫЙ

The stems of certain compound words will become entries in the Glossary due to the device of representing several exceedingly common attributive elements as pseudo-prefixes. (See Appendix II, List of Pseudo-Prefixes)

Examples of such stems are:

БЛАГОДАР

МНОГОЧЛЕН

ПЕРВОБЫТН

СВЕРХПРОВОД

B. Adapted Words

Certain words in the Russian language are termed "International Words" by Russian linguists. Others call them "Adapted Words" or "Non-Slavic Words". The term "Adapted Words" will be used hereafter. These words are usually of Greek or Latin derivation and are used in scientific literature throughout the world. Usually a suffix is added to the root so the word may be inflected. Sometimes the root is prefixed by a purely Slavic element. As a rule the roots of these "adapted" words may be transliterated without any loss of meaning. The occurrence of such words is frequently signalled by the appearance of certain characteristic affixes, as well as the single letters э or ф -- initially or medially--or of А initially. (See Appendix IV.)

PART II. GLOSSARY LOOK-UP

In our current, preliminary scheme, glossary look-up for true Cyrillic words will be performed in the following steps:

A. The complete source word is compared against certain Special Word Lists. Each portion of a hyphenated word is treated as a separate word during the glossary look up only. If a match is not found the routine continues.

B. The reflexive ending, СЬ or СЯ (if any) is removed and stored in a specific location for the source word.

C. The inflectional pseudo-ending (if any) is removed and stored as in B. (See Appendix I, List of Pseudo-Endings).

D. Each pseudo-prefix possessed by the source word is replaced by a single character. (See Appendix II, List of Pseudo-Prefixes)

E. The remaining characters, if any (see Note 1), are treated as follows: The machine compares a successively diminished number of characters from left to right with the entries in the list of pseudo-roots, internally stored. If a match is found the routine goes to the next step. If a match is not found, the routine considers first that a match has been made with the null pseudo root. Since some pseudo-roots are the same as certain pseudo-suffixes, a stem with an apparent pseudo-root may actually turn out to have a null root; a special subroutine will resolve these cases.

Note 1. If a source stem is composed only of pseudo-affixes, the word is considered to possess a null pseudo-root. An example of such a stem is ВО-ПРО-С.

F. The remaining characters are now compared against a List of Pseudo-Suffixes. If the search is successful all characters will be exhausted. Each pseudo-suffix is replaced by a single character and the pseudo-root, including the null root, is replaced by an address referring to its location in the External Memory. This highly compacted representation of a stem is called a Transform. Thus each stem is represented uniquely by its transform. This device saves storage space and reduces sorting time; it has no other significance.

G. If in F above not all characters are exhausted, the source word is considered to be either an adapted word or a compound word. In such cases the machine is directed to a subroutine which proceeds as follows:

1. The stem is transliterated and its identification tag is appended. (cf. 1/)

2. In the residue remaining in F, the first character is tested for a linking vowel o or e . If it is one of these and a residue still remains, we consider the treated portion, temporarily, as a Slavic attribute of a compound word and the still unmatched residue is treated as another possible Slavic stem.

Since this takes place internally the final decision as to the correctness of the decomposition rests on two factors:

a. That the transform is found in the External Glossary. The routine does not have immediate access to the External Glossary. In order to curtail expenditure of time and money the External Glossary is only consulted after the internal sorting file has been filled.

b. That the transform is marked as being a possible element of a compound word. (Most words will not be elements of compound words.) This marking will resolve many cases of incorrect decomposition of a source word. It may be noted here that the translation of a word used as an element of a compound word may be different from its translation when it is used independently. The word МЕЖДУ standing alone is translated "among/between", but as an attributive element it will have the target "inter/ -- ".

Illustration

Examples of such occurrences are:

Example 1. ВЕРЕТЕНООБРАЗНЫЙ

a) The ending ЫЙ is removed and stored.

b) The pseudo-prefix В is replaced by a special character.

c) The pseudo-root ЕРЕТ is identified.

d) The pseudo-suffix ЕН is replaced by a single character.

e) Since no further pseudo-suffixes can be identified, the next character is tested and is found to be the linking vowel О .

f) The process is repeated for the residue ОБРАЗН . The machine identifies ОБ and РАЗ as pseudo-prefixes. The scheme then enters the special subroutine mentioned in E, since the remaining character Н is identified as being both a pseudo-root and a pseudo-suffix. The subroutine determines that Н is a pseudo-suffix since no other characters remain in the residue. Thus, the base element in this compound word possesses a null pseudo-root.

In this example, both elements will be verified as Slavic by the Glossary, and the first element will be marked as a possible attribute.

Example 2.

ВЫСОКОТЕМПЕРАТУРНЫЙ

Having temporarily identified, as in the previous example, the portion ВЫСОК as a Slavic attribute followed by the linking element О , we attempt to treat the residue as if it were also a Slavic stem. The pseudo-root ТЕМ is found, but no succeeding portion of the residue can be identified as a pseudo-suffix. Subsequent search in the Glossary will verify the facts that the first element of the word should be translated and the second element transliterated.

3. If a linking vowel does not occur, or if no match can be found with pseudo-roots and/or pseudo-suffixes, successively diminished portions of the entire stem are compared once more from left to right against lists of common Slavic and non-Slavic attributive elements. (See Appendices III and IV)

If a match is found, the remainder of the word is treated as a new stem.

EXAMPLES ARE:

a. МЕЖДУНАРОДН-ЫЙ

1) МЕЖДУ is found in the list of common Slavic attributes. It will eventually be translated.

2) The remainder of the stem decomposes into НА-РОД-Н . The transform will be found in the External Stem

Glossary and it too will be translated. The first attribute will have one of its several targets, namely "inter", marked as a suitable translation when appearing in combination with other stems. Similarly, in the second word, the target "national" will be thus marked. The entire source word will therefore be translated as "inter-national".

b. АНТИБИОТИК-И

1) АНТИ is found in the list of common non-Slavic attributes. It will be transliterated.

2) The remainder of the stem does not decompose according to our scheme. It will be transliterated. The source word will then yield "[antibiotik]-s".

c. АНТИВОЕНН-ЫЙ

1) АНТИ is found in the list of common non-Slavic attributes. It will be transliterated.

2) The remainder of the stem decomposes into ВО-ЕНН . The transform will be found in the External Stem Glossary and it will therefore be translated. The source word will yield, finally: "[anti] military".

4. If a stem begins with a non-Slavic attributive element that is not listed, the above routine is not sufficient. In this case scanning will begin from right to left in an attempt to identify the base element as a Slavic stem. If it succeeds the next character to the left is examined for a linking vowel. If that be the case the process continues, since the

word may have more than one attributive element. The identified portion will be translated. The unidentified portion will be transliterated.

An example of such a word is ЦЕНТРОСТРЕМИТЕЛЬНЫЙ. The routine when scanning from left to right could not identify the initial portion. Scanning from right to left, however, revealed a true Slavic word СТРЕМИТЕЛЬНЫЙ preceded by a linking vowel. The identified portion will be translated while the remainder of the word will be transliterated.

As we have indicated above, the seemingly successful decomposition of a stem, or a portion of a stem, is not a proof of its correctness.

An example of a false decomposition is the word РЕВОЛЬВЕР. This word would be decomposed into an attributive element РЕВ followed by a linking vowel О, another attributive element ЛЬВ with the linking vowel Е and the base element Р. In the final search of the Stem Glossary, no stem Р would be found; nor would either of the other stems be marked as possible attributes. Therefore, the entire stem would appear transliterated in the Print-out.

In a similar manner the decomposition of ПРОТОКОЛ would prove to be incorrect. It would remain transliterated.

5. Whenever scannings of a stem in both directions fail to identify it as a compound stem, the routine relinquishes its effort at further precision, since it may be assumed that the unidentified stem is one of the following:

a. An adapted word.

b. A proper name.

c. A symbolic expression.

d. An error in input.

e. An omission in our Glossary.

An adapted word will be handled in the following manner. The "international" root and affixes (if any) will be transliterated. The Russian affixes will be translated.

For example:

ТЕОРЕТИЧЕСКИЙ

1. The ending ИЙ is removed and stored.

2. The Russian suffix ИЧЕСК is translated "ICAL".

3. The international root ТЕОРЕТ is transliterated.

RESULT: [TEORET] - ICAL

The necessary morphological information will be derived from the suffix and the ending.

Proper names and symbolic expressions (in Cyrillic characters) of course remain transliterated. Errors in input and Glossary omissions will be traced by means of their identification tags, and corrective measures will be applied.

APPENDIX I

List of Pseudo-Endings.

А	ИЕ	ИЯМ	УЮ
АМ	ИЕВ	ИЯМИ	ШИ
АМИ	ИЕЙ	ИЯХ	Ы
АТ	ИЕМ	Й	ЫЕ
АХ	ИЕЮ	ЙТЕ	ЫЙ
АЯ	ИИ	Л	ЫМ
В	ИЙ	ЛА	ЫМИ
Е	ИЛ	ЛИ	ЫХ
ЕВ	ИЛА	ЛО	Ь
ЕГО	ИЛИ	О	ЬТЕ
ЕЕ	ИЛО	ОВ	ЬЮ
ЕЙ	ИМ	ОГО	Ю
ЕМ	ИМИ	ОЕ	ЮТ
ЕМУ	ИТ	ОЙ	ЮЮ
ЕТ	ИТЕ	ОМ	Я
ЕТЕ	ИТЬ	ОМУ	ЯМ
ЕШЬ	ИХ	ОЮ	ЯМИ
ЕЮ	ИШЬ	ТЬ	ЯТ
И	ИЮ	У	ЯХ
ИВ	ИЯ	УТ	∅ = Null

APPENDIX II.

LIST OF PSEUDO-PREFIXES

(arranged by length, alphabetically)

В*	ОБ*	ПРИ*	ДОЛГО
О*	ОС	ПРО*	КРИВО
С*	ОТ*	РАЗ*	МНОГО
У*	ПО*	РАС	ПЕРВО
Х	СО*	ТРЕ	ПРЯМО
ъ	БЕЗ*	ТРИ*	РАВНО
ВЗ	БЕС	ДВОЕ*	РАЗНО
ВО*	ВСЕ*	МАЛО	СВЕРХ*
ВЫ*	ВОЗ	ОБЩЕ	ШЕСТИ
ДО*	ВОС	ОДНО	ВОСЬМИ
ЗА*	ВНЕ*	ПЕРЕ	ДЕВЯТИ
ИЗ*	ДВУ	ПРЕД*	ДЕСЯТИ
ИС	НАД*	ПЯТИ	КРАТКО
НА*	НАИ	САМО	ПРАВДО
НЕ*	ПОД*	СВОЕ	ЧЕТЫРЕ
НИ*	ПРЕ	БЛАГО	ПРОТИВО

* Also in Special Word List

APPENDIX III

The following additional entries in the Special Word List not already noted in Appendix II may be elements of compound words.

КОЕ

ЛИБО

МЕЖ

МЕЖДУ

НИБУДЬ

НИЖЕ

ПОЛ

ПОЛУ

ПОСЛЕ

ПРЕЖДЕ

ТАКИ

ТО

APPENDIX IV

The following non-Slavic prefixes may be elements of compound words.

АВТО	ИНТЕР	МУЛЬТИ
АГРО	ИНТРА	НИЛЬ
АНТИ	КВАЗИ	ПСЕВДО
АЭРО	КИЛО	СУБ
БИО	КОНТР	СУПЕР
ГЕТЕРО	МАКРО	ТЕЛЕ
ГИДРО	МЕГА	ТЕРМО
ГИПЕР	МЕЗО	УЛЬТРА
ГИПО	МЕТА	ФОТО
ГОМО	МИКРО	ЭКВИ
ИЗО *	МИЛЛИ	ЭКСТРА
ИНФРА	МОНО	

* This is also a combination of the two Slavic prefixes ИЗ and О.

References

1. Ida Rhodes, A New Approach to the Mechanical Syntactic Analysis of Russian, MT VI, 2, 1961.

2. Н. Г. Хромец, Учебник Русского Языка Для Нерусских, , Moscow 1959.

3. С. Г. Бархударов, Учебник Русского Языка, Moscow 1960.

4. В. В. Виноградов, Грамматика Русского Языка, Academy of Sciences, Moscow 1960.

U. S. DEPARTMENT OF COMMERCE
Luther H. Hodges, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its major laboratories in Washington, D.C., and Boulder, Colorado, is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside of the front cover.

WASHINGTON, D.C.

Electricity. Resistance and Reactance. Electrochemistry. Electrical Instruments. Magnetic Measurements. Dielectrics.

Metrology. Photometry and Colorimetry. Refractometry. Photographic Research. Length. Engineering Metrology. Mass and Scale. Volumetry and Densimetry.

Heat. Temperature Physics. Heat Measurements. Cryogenic Physics. Equation of State. Statistical Physics.

Radiation Physics. X-ray. Radioactivity. Radiation Theory. High Energy Radiation. Radiological Equipment. Nucleonic Instrumentation. Neutron Physics.

Analytical and Inorganic Chemistry. Pure Substances. Spectrochemistry. Solution Chemistry. Analytical Chemistry. Inorganic Chemistry.

Mechanics. Sound. Pressure and Vacuum. Fluid Mechanics. Engineering Mechanics. Rheology. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Crystal Growth. Physical Properties. Constitution and Microstructure.

Building Research. Structural Engineering. Fire Research. Mechanical Systems. Organic Building Materials. Codes and Safety Standards. Heat Transfer. Inorganic Building Materials.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Applications Engineering.

Atomic Physics. Spectroscopy. Radiometry. Solid State Physics. Electron Physics. Atomic Physics.

Instrumentation. Engineering Electronics. Electron Devices. Electronic Instrumentation. Mechanical Instruments. Basic Instrumentation.

Physical Chemistry. Thermochemistry. Surface Chemistry. Organic Chemistry. Molecular Spectroscopy. Molecular Kinetics. Mass Spectrometry. Molecular Structure and Radiation Chemistry.

• Office of Weights and Measures.

BOULDER, COLO.

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Ionosphere Research and Propagation. Low Frequency and Very Low Frequency Research. Ionosphere Research. Prediction Services. Sun-Earth Relationships. Field Engineering. Radio Warning Services.

Radio Propagation Engineering. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmosphere Physics.

Radio Standards. High Frequency Electrical Standards. Radio Broadcast Service. Radio and Microwave Materials. Atomic Frequency and Time Interval Standards. Electronic Calibration Center. Millimeter-Wave Research. Microwave Circuit Standards.

Radio Systems. High Frequency and Very High Frequency Research. Modulation Research. Antenna Research. Navigation Systems. Space Telecommunications.

Upper Atmosphere and Space Physics. Upper Atmosphere and Plasma Physics. Ionosphere and Exosphere Scatter. Airglow and Aurora. Ionospheric Radio Astronomy.


NBS